

マンガからの自動人物検出と識別に関する一検討

石井大祐[†] (正会員) 渡辺 裕[†] (正会員)[†]早稲田大学大学院国際情報通信研究科

A Study on Automatic Character Detection and Recognition from Comics

Daisuke ISHII[†] (Member), Hiroshi WATANABE[†] (Member)[†]Graduate School of Information and Telecommunication Studies, WASEDA University

〈あらまし〉近年、マンガは日本国内のみならず海外においても多数の読者を持つコンテンツとなっている。また、電子書籍閲覧可能な端末の進化と電子書籍市場の発展により、電子化されたマンガの需要も高まっている。マンガは絵、文字、擬音などの多数の要素がすべて一枚の誌面上に描かれるため、大変複雑な構造を持つ。これら登場人物や多数の要素はメタデータとして取り出すことで、デジタル化されたマンガの利便性を向上させることができる。マンガは基本的に白黒で描かれるため、画像の特性は自然画像と異なる。このため、マンガの登場人物解析を行うためには、マンガの登場人物の特徴をうまく捉える手法を用いる必要がある。本稿では、HOG (Histograms of Oriented Gradient) と SVM (Support Vector Machine) による登場人物の瞳、顔の検出処理と、登場人物の識別処理を組み合わせることで、マンガのページ上から、特定登場人物の検出を行う手法を提案する。結果として、提案手法による特定登場人物の検出で最大0.93のPrecisionを得た。

キーワード：顔検出、顔識別、マンガ、メタデータ

〈Summary〉 Recently, comics has been contents with many user not only Japan but also foreign countries. Moreover, demand for digital comic is increased by digital book equipment developing and growth of e-book market. Comics have very complex construction because a lot of elements that picture, texts, onomatopoeia, and so on are drawn on a plane. To extract elements and characters as metadata, we can increase usability of digitized comics. Comics generally have been drawn white and black, so the feature of image is different with natural picture. Thus, to analyze comic character, we must select method that is able to treat comic image feature. In this paper, we propose specified character detection methods on a comic page by eye detection, face detection, and person recognition that utilize HOG (Histograms of Oriented Gradient) and SVM (Support Vector Machine). Finally, we obtained 0.93 maximum precision for specified character detection by proposed method.

Keywords: face detection, face recognition, comic, metadata

1. はじめに

近年、電子書籍の発展は著しく、各社より閲覧用デバイスが発売されており、International Digital Publishing Forum (IDPF)によりEPUB 3.0¹⁾が策定されるなど、今後その普及が更に加速すると予測される。電子書籍はこれまでの紙媒体の書籍と異なり、電子的に書籍を購入可能であり、また、技術的に内部コンテンツの検索などによる、コンテンツ情報ベースのアクセスが可能である。

書籍の一ジャンルとして、マンガは現在日本国内だけでなく、海外においても多数のファンを持つ、日本を代表するコンテンツの1つとなっている。また電子書籍市場においてもその存在は大きく、日本における電子書籍市場の発展を牽引してきた²⁾。

電子書籍としてのマンガには、単に紙媒体のマンガを電子化したものだけでなく、様々な付加要素を含むリッチコンテンツとして作成された作品も存在する。具体的には、効果音

や一部画像が振動するなどの視覚効果が付与された作品、一部がアニメーション化された作品などがある³⁾。また、携帯電話向けサービスとして、コマごとに順番に表示する方法や、ページ内部の読み順に沿って順番にスクロールを行う機能が重畳されているものもある。一般的に、電子書籍購入時には作品名や作者、掲載誌などから目的の書籍を検索する事が可能である。一方で、マンガ配信における前述のサービスやリッチコンテンツ作成時に使用されるメタデータの多くは、人手により付与されている。

マンガのアーカイブやコンテンツの利便性を向上させるために、マンガ画像の要素に対する自動解析を実現することは有意である。とりわけ、登場人物に関する情報はマンガの構成上、大変重要である。我々はマンガ内部の情報に基づくメタデータの解析及びその付与の実現を目指し、マンガの解析手法について検討を行っている。マンガは様々な絵やセリフなどをはじめとする要素群の組み合わせで構成されており、その構造は複雑であるため解析が困難である。

マンガは基本的に白地の紙の上に黒色のインク及びトーンを用いて描かれており、電子化されたマンガは本来白黒2値のデータとなる。低解像度化や、入力時のローパスフィルタの影響により一部中間調が出てくるが、全体としてはほぼ白黒2値のデータであると考えることができる。

これまでに顔検出や人物、物体検出に利用可能な様々な手法が検討されてきた^{4)~6)}。また、上記より得られる画像特徴量を学習するために統計的機械学習手法⁷⁾との組み合わせも提案されている。特に、人物については、その濃淡や色情報をうまく活用することで、その検出及び認識が実現されている。

一方で、マンガは2値データの線や点群によって構成されるものであり、実画像に対して有効な上記解析手法をそのまま適用してもうまく解析が行うことができない。また、2値画像からの情報抽出は、OCRや文書解析などの検討がなされてきているが、一般的な文書画像と異なり、マンガは複雑なレイアウトによって構成されるため、十分な内容解析を行うことが困難である。

マンガに対する研究としてこれまでに、マンガ上の付加的情報の一部について、その解析手法が検討されている。帆足らはマンガのシーン要約を目的とした研究の中でコマ及び文字とその吹き出し部分の検出を行なっている⁸⁾。また、野中らはマンガの電子配信向けのオーサリングツールとして、コマの分割について提案を行っている⁹⁾。マンガ画像からの顔や瞳の検出手法^{10)~12)}も検討されているが、その検出条件が厳しいなど、未だ発展途上である。また、アニメ作品を対象として、登場人物識別が検討されている¹³⁾。この手法では、色情報を利用しており、色のないマンガの登場人物識別には利用不可能である。

本稿では、マンガ上のメタデータとして、マンガの登場人物の情報を得ることを目的とし、マンガ画像からの自動人物検出及びその識別手法について検討を行う。マンガの登場人物には、表現による登場人物の表情変化、オクルージョン、描画サイズによる表現の変化等があるため、顔をそのまま学習、検出することは困難である。また、顔の識別に際して、学習に利用する画像の選別方法によりその結果が大きく変動する。そこで、登場人物顔の検出及び識別に対して、Histograms of Oriented Gradient (HOG) 及び Support Vector Machine (SVM) を用い、判別処理と学習画像の組み合わせについて検討を行い、その様々なパターンについて実験によりその識別結果を明らかにする。

本稿の構成を以下に示す。2章では、マンガの特徴についてまとめ、一般的に利用される画像解析手法及びマンガ関連の解析手法について述べる。3章では、今回提案するマンガ画像に対する登場人物検出と登場人物識別について述べる。4章では、3章で述べた手法を用い、実際にマンガ画像に対して検出及び識別処理を実施した結果をまとめ、各結果について考察を述べる。5章で本稿をまとめる。



図1 マンガ画像例^{11), 16)}
Fig. 1 A sample of comic image^{11), 16)}



図2 マンガ画像上のSIFT特徴量^{11), 16)}
Fig. 2 SIFT feature on comic image^{11), 16)}

2. マンガ画像と画像解析手法

2.1 マンガについて

マンガは、主に白色の紙の上に線画と黒色のドットからなるトーンを用いて描かれる。全体の構造は基本的には線画で描かれ、その上にトーンや塗りつぶしなどの手法により陰影の表現が重ねられる。マンガのレイアウトの特徴としては、コマと呼ばれる枠の中に、キャラクターや風景などの絵の部分と、吹き出し、セリフや効果音などの文字要素が複雑に配置されている。また、ページ上を読むための順序はコマによって定義される。ただし、コマからのキャラクターや文字要素、吹き出しなどはみ出しを許容するため、レイアウトは非常に自由度が高い。マンガ画像の例を図1に示す。

2.2 画像解析技術

これまでに顔検出を主な目的として、様々な画像特徴量やその解析手法が検討されてきた。中でも代表的な手法として、ViolaとJonesによる顔の検出手法やScale-Invariant Feature Transform (SIFT) 特徴量がある。また、画像のエッジを考慮した特徴量として、HOG特徴量がある。ここではこれらの手法について概略を示し、マンガ画像への適用に対する考察を述べる。

2.2.1 ViolaとJonesによる顔検出

顔検出を主な目的とした研究成果として、ViolaとJoensはHaar-like特徴量を利用した高速かつ頑健な顔検出手法を提案している⁴⁾。人間の顔は、瞳部分や唇など、多くの共通した濃度的特徴がある。画像に対してHaar-like特徴量を様々なパターンで計算することで、この顔の特徴をうまく捉えることに成功している。

マンガ画像は主に線画により主要成分が描かれている。また、もともとほぼ2値のデータであるため、Haar-like特徴量を計算すると、線画の書き込み量やドットの頻度が得られるが、線の方向等に関する情報は得られない。登場人物の識別過程において、顔のバランスは登場人物間では似た状態であ

る可能性が高く、解析に十分な情報が得られないと考えられる。

2.2.2 SIFT

スケール変化および回転にロバストな画像特徴量として SIFT 特徴量⁵⁾が提案されており、近年画像の特定物体認識や、画像の照合などに広く用いられている。SIFT は、Difference of Gaussian (DoG) を用いてスケール不変なキーポイントを検出し、SIFT descriptor により回転不変な特徴量を記述する手法である。そのため SIFT 特徴量は、回転・スケール変化にロバストな特徴量となっている。

一方でマンガ画像では、線画、トーン部分の双方において、SIFT のキーポイントになりうるエッジ成分が無数に存在する。このため、画像中のあらゆる箇所に SIFT のキーポイントが検出されてしまう。これは、完全に同一の画像や部位を照合する目的には使用できるが、変化のある特定の登場人物などを見つけ出す用途にはキーポイントの制御が困難であり、また、多数のキーポイントから有意な情報のみを取り出すことは難しい。マンガ画像上で SIFT 特徴量取得した結果例を図 2 に示す。無数のキーポイントが検出されている事が確認できる。

2.2.3 HOG

エッジの方向性を捉える特徴量の一つとして、HOG 特徴量⁶⁾が提案されている。HOG 特徴量は、対象領域を複数のセルに分割し、各領域内で方向ヒストグラムを計算する。得られた方向ヒストグラムをまとめて多次元ベクトルを構成し特徴量としている。これにより、画像内部の物体形状を捉えるのに適した画像特徴量となっており、シルエットによる人物検出などに利用されている。

マンガは多数の線の組み合わせを駆使することで登場人物を表現している。したがって、これは人物におけるシルエットの形状を表現していることに近く、方向性の情報を捉えることが可能な HOG 特徴量は、マンガ画像の解析にも有効であると考えられる。

2.3 マンガ画像解析技術

マンガの電子配信支援や、内容解析を目的として、マンガ内の要素解析に関する研究が行われている。現在提案されている研究をその要素から分けると、コマ分割や文字と吹き出し抽出を行うもの^{8),9)}と、登場人物の検出^{10)~12)}を目的としたものがある。

コマ分割では、マンガのページ内部における構造の主構成要素であるコマの位置及びその並びを検出することが目的である。マンガのコマは主に矩形に限られない四角形で、さらに、コマから内部要素がはみ出す場合もある。このため、従来の文書画像向けに提案されてきたレイアウト解析技術ではその解析が困難である。そこで、コマの枠線部分を用いてうまく切り分けることで、コマ分割を行う。また、文字と吹き出しの抽出では、文字部分を検出し、そこを起点として吹き

出しの検出を行う。

マンガ上の登場人物検出に対してこれまでに瞳検出と顔検出に関する検討がなされている。我々はこれまでに、HOG 特徴量と統計的機械学習である SVM を用いた画像上の登場人物の瞳部分の検出を試みている^{10),11)}。実際のマンガ画像では、登場人物の顔の変化やオクルージョンの影響があるため、顔検出の前段階として瞳の検出を行なった。また、新井ら¹²⁾は、同じく HOG 特徴量と SVM を用いた顔の判定を用い、顔の検出範囲を予め制限する前処理を導入した手法を提案している。しかし、人物の方向を制限した場合のみの検出を試みている段階であり、また、人物識別については未検討である。

3. マンガからの自動人物検出及び識別

3.1 概要

本研究の目的は、マンガ画像からの自動による人物検出及び識別の実現である。マンガ上の登場人物はストーリーを構成する上での役者であり、その表情変化は激しく、また、限られたサイズのページ内に様々な要素が多数配置される特性から、オクルージョンが多い。加えて、登場人物は描かれる大きさや向き、表現により離散的に変化する。したがって、画像上から登場人物を直接検出し、識別結果を得ることは困難であると考えられる。そこで、本研究の目的を人物の存在有無の特定、人物の顔位置の特定、人物の識別の 3 ステップに分解し検討を行う。

3.2 瞳、顔、人物識別器

本稿における識別処理で使用する、各識別器について述べる。識別器は我々のこれまでの検討^{10),11)}より、HOG 特徴量とこれを学習、識別するための SVM によって構成することとした。

HOG 特徴量はシフトおよび回転に対してロバストではないため、本稿では画像上における同一箇所の瞳や顔について、複数のパターンで位置や範囲などを変化させて切り出した画像を学習に使用する。マンガ画像上では、様々な大きさで顔や瞳が描かれるが、このスケールの違いに対しては、我々の従来の研究¹¹⁾にて、解像度変換を行っても HOG 特徴量を利用した瞳検出処理結果において、True Positive の数に大きな影響がなくまた、False Positive の数が少ない結果が得られている。そこで、本稿では、学習用及びスライドウィンドウにより切り出された画像に対して、解像度変換により、すべて均一の解像度に正規化し、HOG 特徴量を計算する。

3.3 人物検出及び識別について

人物の存在有無の特定については、3.1 節に述べた理由により、顔全体の安定した検出は困難である場合がある。一方で、瞳はマンガの中において、他の箇所と比較して特徴的な描かれ方をする箇所であり、瞳が検出できれば、そこに人物

がいる可能性が高いと判断できる。このため、顔全体の検出処理だけでなく、瞳部分に絞って検出処理を行い、人物の存在有無を捉えることが考えられる。

次に、人物の顔位置の特定については、顔を学習した検出器を用い、スライドウィンドウを用いた検索を行う方法が考えられる。また、瞳の検出結果が得られている場合には、瞳周辺に対してのみ顔の検出処理を行うことも可能である。

人物の識別については、検出対象者ごとに、当該の検出対象者と、検出対象者以外の顔画像を学習することで、検出対象者であるか否かの識別を行う。

3.4 瞳、顔および人物識別器とその組み合わせ

前節で検討した識別器をまとめると以下の3つとなる。ただし、すべての学習用正例および負例の画像は、あらかじめ手作業にて切り出されたものを使用する。

瞳識別器：瞳部分の画像を正例、瞳以外の部分の画像を負例として学習を行ったもの。

顔識別器：顔部分の画像を正例、顔以外の部分の画像を負例として学習を行ったもの。

人物識別器：検出対象人物の顔を正例、検出対象以外的人物の顔を負例として学習を行ったもの。

筆者らのこれまでの調査により、各識別器は多くの場合弱い識別器となり、多数のFalsePositiveを結果として出力する。そこで本稿では、これらの組み合わせによる瞳検出、顔検出、人物識別処理について提案する。各ステップを考慮した、上記識別器の組み合わせについては以下のパターンが考えられる。

Eye: 画像からスライドウィンドウと瞳識別器を用いて瞳の検出を行う

Face: 画像からスライドウィンドウと顔識別器を用いて顔の検出を行う

Character:

画像からスライドウィンドウと人物識別器を用いて得られた領域が、特定キャラクターの顔であるか識別を行う

Eye-Face(E-F):

Eyeを実施し瞳の候補領域を取得後、その領域の周囲に対して顔識別器を用いて顔検出を行う

Eye-Character(E-C):

Eyeを実施し瞳の候補領域を取得後、その領域に対して人物識別器を用いて特定登場人物の顔であるか識別を行う

Face-Character(F-C):

Faceを実施し、顔の候補領域を取得後、その領域に対して人物識別器を用いて特定登場人物の顔であるか識別を行う

Eye-Face-Character(E-F-C):

Eye-Faceを実施し、顔の候補領域を取得後、その領域に対して人物識別器を用いて特定登場人物の顔であるか識別を行う

ここで、上記の識別に用いる特徴量としてHOG特徴量を利

用する。Eye, Face, Characterでは、入力されたマンガ画像に対してスライドウィンドウを用いて瞳識別器、顔識別器、人物識別器を用いて検出処理を行う。Eye-Faceでは、瞳検出結果の周囲に対して、複数サイズの領域で新たにウィンドウを定義し、このウィンドウに対して顔識別器による判別処理を行う。顔と認められる範囲があれば、当該ウィンドウを顔と判定する。Eye-Character(E-C), Face-Character(F-C)はそれぞれ瞳検出、顔検出の結果に対して顔識別処理を行い、検出対象の登場人物か否かの判定を行う。Eye-Face-Character(E-F-C)ではEye-Faceの結果として顔であると判定されたウィンドウに対して顔識別処理を実施する。

本稿では、マンガ画像における顔の変化度合いを考慮した場合に、瞳の方がその変化パターン総数が少ないと予想されるため、顔の検出ができない場合でも人物の存在を検知するためには、瞳の有無から得られる情報を優先させることが有効であると考えられる。そこで、顔および特定人物検出においても、瞳検出と顔検出の順番については、先に瞳検出を行うこととした。

4. 評価実験

評価実験では、まず人物識別器について、あらかじめ手作業により切り出された顔画像に対する識別実験結果について示す。続いて、3.4節にて提案手法として述べた、各識別の組み合わせについて、実際にマンガ画像に対して実験を行った結果を示す。また、上記組み合わせに加え、一部をHaar-like特徴量を用いて置き換え、HOG特徴量とHaar-like特徴量の比較実験結果も併せて示す。

4.1 実験条件

本稿では、SVMの実装にChangらによるLIBSVM¹⁴⁾を用いた。HOG特徴量は方向の正規化数を9、ブロックサイズを3×3セル、セルのサイズはHOG特徴量計算対象の画像サイズに対して適応的に決定することとし、対象画像を縦方向横方向それぞれ5分割するように設定した。したがってHOG特徴量の次元数は9(ブロック内に含まれるセルの数)×9(正規化回数)×9(方向)=729(次元)となる。比較実験に用いたHaar-like特徴量については、OpenCV¹⁵⁾の実装を使用した。また、Haar-like特徴量の一般的な利用状況を考慮し、学習にはAdaBoostを利用したOpenCVの実装であるtraincascadeを用いた。予備実験より学習用画像のサイズは、瞳、顔検出用は16×16、登場人物識別用は32×32画素とした。

4.2 人物識別器に関する実験

1作品に登場する主要な2人の登場人物A, Bについて、本人を正例、本人以外の登場人物を負例とし、あらかじめ切り出された顔画像群に対して識別実験を行った。学習に使用した画像枚数を表1に、結果を表2にそれぞれ示す。A2は登場人物Aに対する学習用正例画像の切り出し元となる画像数を

表1 顔識別の学習と実験に用いた画像数

Table 1 The number of images utilized for learn and examination of face recognition

Data Set Character	Known		Unknown	
	Positive	Negative	Positive	Negative
A	93	32	92	86
A2	36	32	92	86
B	26	45	19	142

表2 顔識別成功率

Table 2 Face recognition success rate

Method	Character	Known		F
		Precision	Recall	
HOG	A	0.94	1.00	0.97
HOG	A2	0.95	1.00	0.97
HOG	B	1.00	1.00	1.00
Haar-like	A	0.76	1.00	0.87
Haar-like	A2	0.61	0.97	0.75
Haar-like	B	0.75	0.92	0.83

Method	Character	Unknown		F
		Precision	Recall	
HOG	A	0.57	0.99	0.72
HOG	A2	0.58	0.92	0.71
HOG	B	0.55	0.58	0.56
Haar-like	A	0.56	0.85	0.68
Haar-like	A2	0.52	0.39	0.45
Haar-like	B	0.28	0.53	0.36

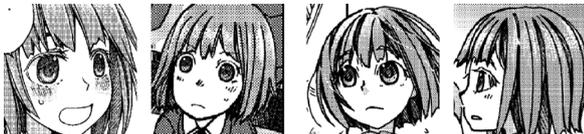


図3 登場人物Aの正例画像

Fig. 3 Positive sample of character A



図4 登場人物Bの正例画像

Fig. 4 Positive sample of character B



図5 瞳の正例画像

Fig. 5 Positive sample images for eye learning



図6 顔の学習用正例画像

Fig. 6 Positive sample images for face learning



図7 瞳および顔の学習における負例画像

Fig. 7 Negative sample images for eye and face learning

表3 既知画像に対する検出結果

Table 3 Detection results for known image

Method	Correct	Incorrect	Total	Precision
Eye	109	406	515	0.211
Face	536	551	1087	0.493
Character(A)	6935	40574	47509	0.146
Character(B)	1104	13335	14439	0.076
E-F	34	5	39	0.872
E-C(A)	149	682	831	0.179
E-C(B)	64	86	150	0.427
F-C(A)	222	498	720	0.308
F-C(B)	117	204	321	0.364
E-F-C(A)	4	9	13	0.307
E-F-C(B)	23	1	24	0.958
HaE	43	280	323	0.133
HaF	89	203	292	0.305
HaF-C(A)	111	475	586	0.189
HaF-C(B)	10	11	21	0.476
HaF-HaC(A)	29	156	185	0.157
HaF-HaC(B)	22	163	185	0.119
E-F-HaC(A)	4	35	39	0.103
E-F-HaC(B)	27	11	38	0.711

表4 未知画像に対する検出結果

Table 4 Detection results for unknown image

Method	Correct	Incorrect	Total	Precision
Eye	157	690	847	0.185
Face	935	928	1863	0.502
Character(A)	12560	66778	79338	0.158
Character(B)	1368	21707	23075	0.059
E-F	44	23	67	0.657
E-C(A)	365	1154	1519	0.240
E-C(B)	42	136	178	0.236
F-C(A)	420	851	1271	0.330
F-C(B)	151	366	517	0.292
E-F-C(A)	22	25	47	0.468
E-F-C(B)	15	1	16	0.938
HaE	70	482	552	0.127
HaF	152	349	501	0.303
HaF-C(A)	239	747	986	0.242
HaF-C(B)	15	23	38	0.395
HaF-HaC(A)	59	216	275	0.215
HaF-HaC(B)	33	242	275	0.120
E-F-HaC(A)	24	43	67	0.358
E-F-HaC(B)	19	44	63	0.302

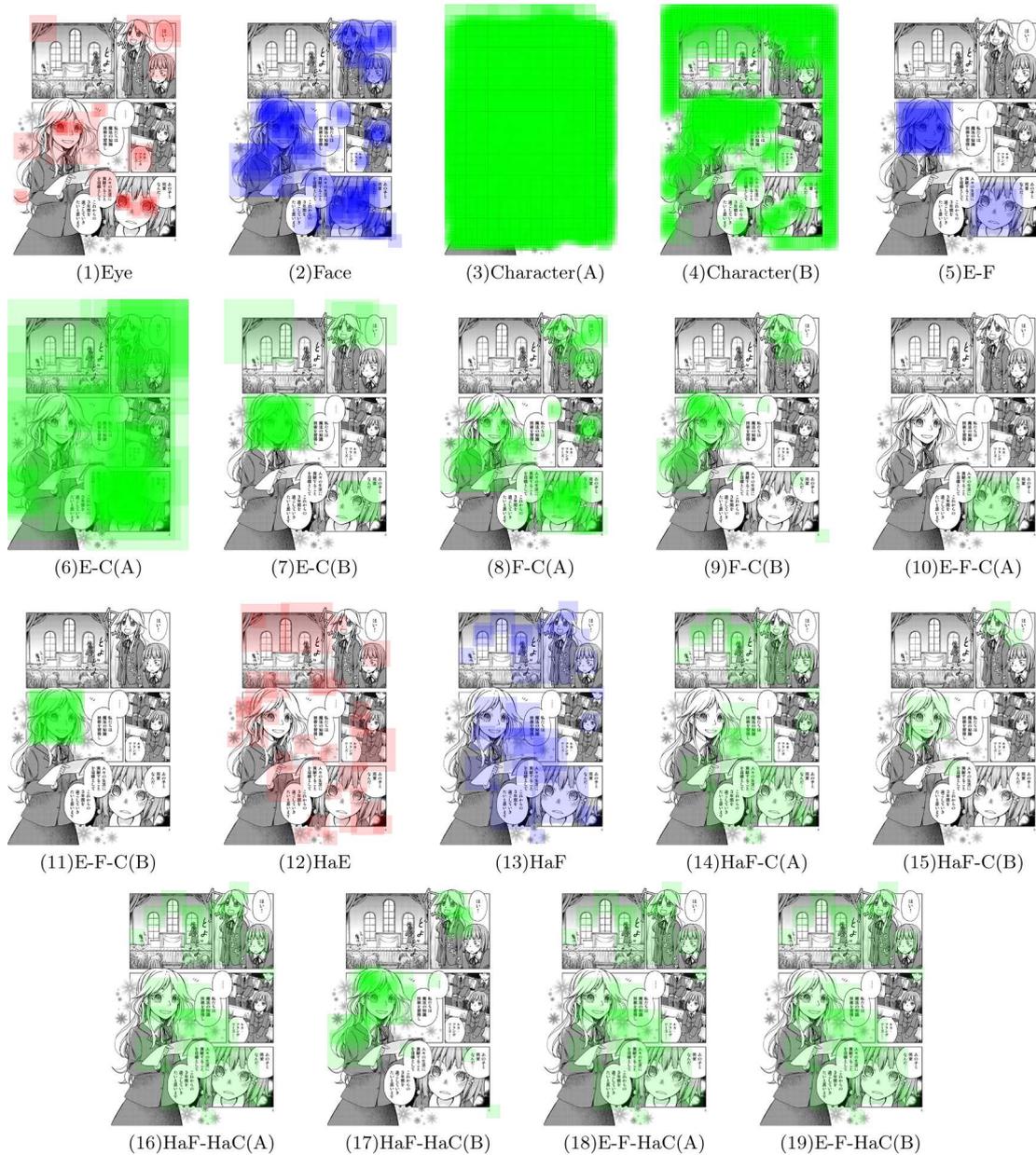


図8 既知画像上の瞳、顔検出及び顔識別結果

Fig. 8 Results of eye detection, face detection and face recognition on known image

制限し、学習枚数を減らしたものである。学習に使用した画像に対する識別結果をKnown、学習に使用していない未知の画像に対する識別結果をUnknownにそれぞれ示す。ここで、学習及び識別対象の顔画像については、様々な方向を含む顔画像群を使用した。評価にはPrecisionとRecall及びその平均値であるF値を利用した。ここで $Precision = (\text{検出された正解数}) / (\text{全検出結果数})$ 、 $Recall = (\text{検出された正解数}) / (\text{全正解数})$ 、 $F \text{ 値} = (2 \times Precision \times Recall) / (Precision + Recall)$ とした。また、検出された正解数は検出対象人物の画像に対してPositiveと判定した数、全検出結果数は、本人・他人を問わずPositiveと判定された数、全正解数は実験対象画像群全体に含まれる、検出対象人物画像の総数である。各登場人物の学習に使用した登場人物AおよびBの正例画像の一部

を、図3及び図4にそれぞれ示す。

4.3 瞳、人物及び特定人物検出実験

各識別器の組み合わせによる検出実験を行い、その結果を示す。ここでは、識別器の組み合わせとして3.4節に示したHOG特徴量を用いた組み合わせに対して、下記のHaar-like特徴量を用いた検出、識別処理を追加し比較を行う。

Eye(HaE): 画像からスライドウィンドウとHaar-like特徴による瞳識別器を用いて瞳の検出を行う

Face(HaF): 画像からスライドウィンドウとHaar-like特徴による顔識別器を用いて顔の検出を行う

Face-Character(HaF-C):

(HaF)を実施し、顔の候補領域を取得後、その領域に対し

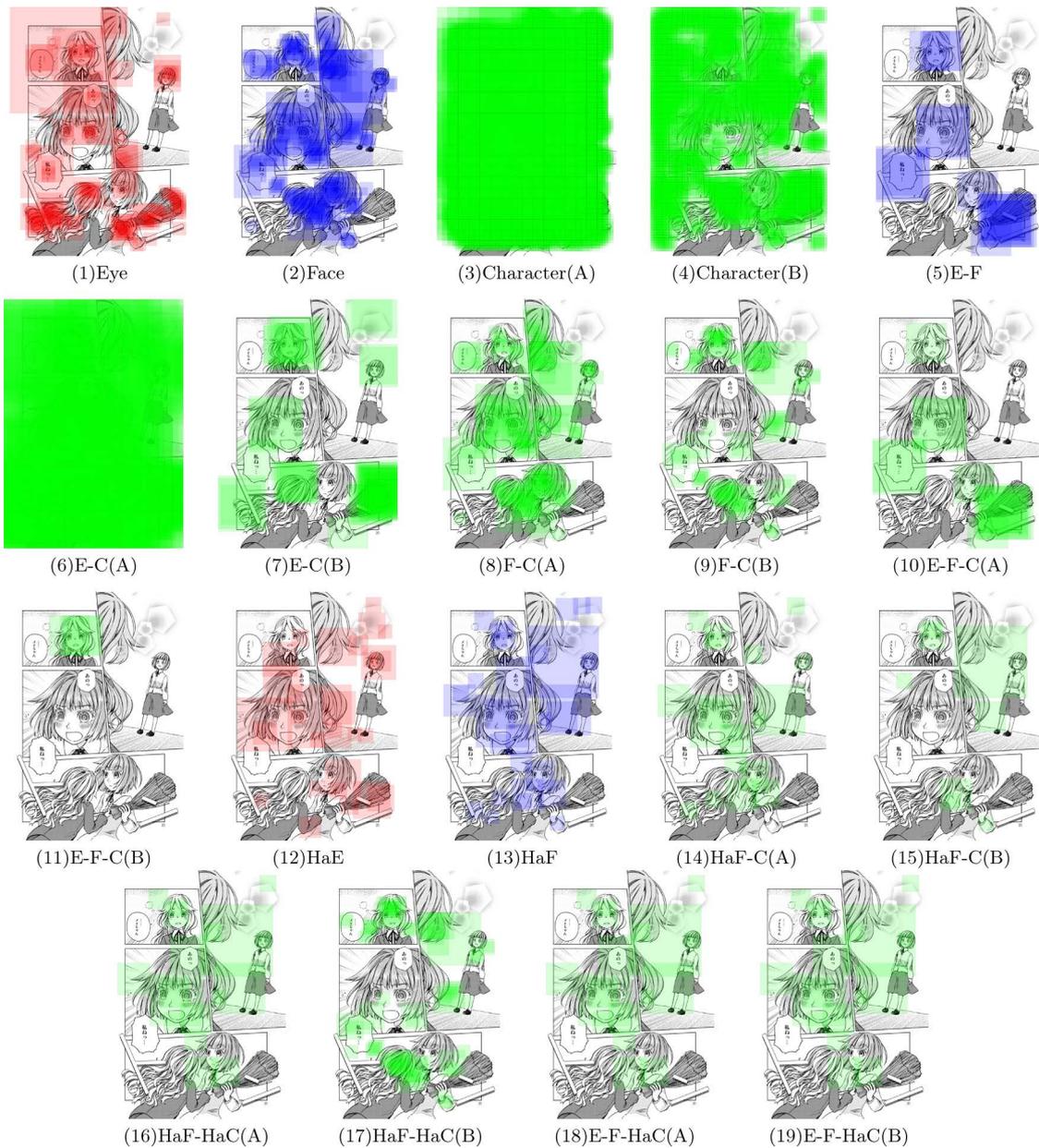


図9 未知画像上の瞳, 顔検出及び顔識別結果

Fig. 9 Results of eye detection, face detection and face recognition on unknown image

てHOG特徴による人物識別器を用いて特定登場人物の顔であるか識別を行う

Face-Character(HaF-HaC):

(HaF)を実施し, 顔の候補領域を取得後, その領域に対してHaar-like特徴による人物識別器を用いて特定登場人物の顔であるか識別を行う

Eye-Face-Character(E-F-HaC):

(E-F)を実施し, 顔の候補領域をHOG特徴により取得後, その領域に対してHaar-like特徴による人物識別器を用いて特定の登場人物の顔であるか識別を行う

登場人物(A, B)に対する識別器の学習に用いた画像及び画像数は前節で示したものと同一である. 実験で用いた入力画像の解像度は1342×1877画素, スライドウィンドウのサイ

ズは50×50, 100×100, 200×200, 400×400画素とし, スライドウィンドウのスライド幅は, ウィンドウサイズの1/5とした. 瞳検出後に実行される顔識別処理のウィンドウサイズは検出された瞳のウィンドウサイズの3倍, 4倍, 5倍とした. 各ウィンドウによって得られた領域はすべて100×100画素に正規化を行い, HOG特徴量の計算を行うものとする. スライドウィンドウのサイズ及び正規化解像度などは, 入力画像のサイズに応じて適宜変更する必要があるが, 予備実験から, 今回の入力画像に対して良好な結果の得られた値を使用した. 今回は, 合計27枚の画像に対して実験を実施した.

瞳検出器の学習に用いた正例画像の一部を図5に, 顔検出器の学習に用いた正例画像の一部を図6に, 双方に用いた負例画像の一部を図7にそれぞれ示す.

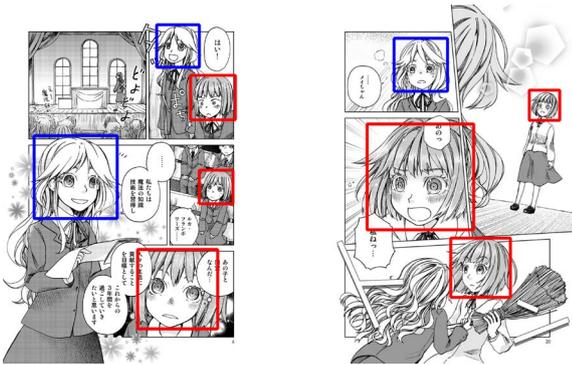


図10 図8及び図9における登場人物A, Bの正解領域例
Fig. 10 Correct resion of character A and B on Fig. 8 and Fig. 9

上記各組み合わせに対する識別結果として、学習に利用した画像を含む既知の画像に対する識別結果を表3に、学習に利用した画像を含まない未知の画像に対する識別結果を表4にそれぞれ示す。今回検出対象画像上におけるすべての検出対象を記述したデータが用意できなかった。このためRecallについてはこれを求めず、Precisionのみの評価とした。Precisionの評価基準として、Eyeに対しては得られた領域が主に瞳部分を含む場合、FaceおよびE-Fについては得られた領域が主に顔部分を含む場合をCorrectとした。また、特定人物検出については、対応する各登場人物の顔を含む場合のみをCorrectとして評価を行った。上記に含まれない全ての検出結果はIncorrectとしてカウントし、Precisionは $Precision = \frac{Correct}{Correct + Incorrect}$ とした。

次に、学習に利用した既知画像に対する検出結果例を図8に、未知画像に対する検出結果例を図9にそれぞれ示す。図中にて赤色でプロットされた領域は瞳の検出結果を、青色にてプロットされた領域は顔の検出結果を、緑色にてプロットされた領域は登場人物の検出結果をそれぞれ示す。括弧の中のAおよびA2とBは4.2節にて示した登場人物A, Bに対応する。また、上記画像例に対する登場人物A, Bの正解位置を図10に示す。ここで、赤枠で示された箇所が登場人物A、青枠が登場人物Bの位置を表す。

4.4 考察

はじめに、人物識別器に関する実験について考察を述べる。識別結果より、Knownに対する識別結果は高く、学習自体は行えていると考えられる。その上で、Unknownに対する識別結果において、評価尺度であるF値がHOGの場合には0.56から0.72であり、未知の画像に対する識別も実施可能であると考えられる。また、Haar-like特徴を利用した場合も、登場人物Aに対してのF値が0.68であり、うまく特徴を捉えられる場合もあると考えられる。

今回学習数による検出結果への影響は軽微であった。ただし、他の予備実験ではこの学習枚数内での変動が大きい場合には、検出結果に対して悪影響を与えることが確認されてお

り、学習用データセットの作成においては、普遍的なシーンから画像を集めることが有効であると考えられる。

次に、瞳、人物及び特定人物検出実験についての考察を行う。実験結果の表より、特定人物検出を行う場合には、HOGを用いたE-F-Cのパターンが最も識別成功率が高くなる結果が得られた。人物の顔同士を学習したCharacterでは、得られている識別器が人物とそれ以外を見分けるにあたって非常に弱い。図8, 9の(3), (4)からも、ほぼすべての領域でTrueの判定を行っていることが確認される。

2種以上の識別器を用いた場合、E-Fは顔をそのまま検出した結果であるFaceよりも良い結果が得られた。一方で、EもしくはFと組み合わせた場合の特定人物のPrecisionは0.24から0.33とE-F-Cと比べ低い値である。E-F-Cの結果はいずれの場合においても他の組み合わせに比べて良好であり、これは、E-Fの結果を利用することで、より高い精度で得られた顔候補領域に対する特定人物の識別処理が実施できたためと考えられる。したがって検出処理をカスケード化することは有効であるといえる。

HOGを利用した場合とHaar-likeを利用した場合の比較では、既知画像に対する一部と、未知画像に対して直接登場人物の顔を得ようとする場合にのみHaar-likeがやや良好な結果である。しかし、実験結果より今回対象としている瞳、人物、特定人物の検出に対しては、HOGを用いた場合のほうが有効であることが確認された。

未知画像と既知画像に関しては、手で切り出した顔画像を用いた識別実験においては顕著な差が見られたが、実際の画像に対する検出を伴う実験においては、その差はほぼないと判断される。これは、スライドウィンドウによる走査が処理の根本にあるため、この移動幅により、実際に判別対象となっている画像と、学習に使用した画像が一致しないことが原因として挙げられる。

画像上の評価としては、図8の(10), (11)にて、それぞれ検出対象の登場人物A, Bを取得できており、検出対象外の登場人物は検出結果から除外されていることが確認される。図9の未知画像に対する実験結果において、登場人物Aを対象とした検出処理では、(11)では検出対象外の人物を検出結果から除外することに成功している。

5. むすび

本稿では、マンガ画像の解析として、マンガ画像上の登場人物の自動検出、及び特定登場人物の検出という課題に対して、マンガの瞳検出、顔検出、登場人物の顔識別処理を組み合わせによる特定登場人物の自動検出処理を提案した。各処理の検出性能及び各処理の組み合わせによる検出精度を実験により明らかにした。結果として、瞳検出、顔検出、登場人物の識別処理すべてを組み合わせの場合に、未知画像に対して最大0.93のPrecisionで検出を行うことができた。今後の検討課題として、画像上の登場人物検出における再現率等につ

いて調査が必要である。

謝 辞

本稿にて例示したマンガ画像はすべて木野陽様<http://www.etheric-f.com/>より学術目的のために使用許可をいただいたものである。マンガ画像の提供及び原稿への掲載を許可いただいた木野陽様に深く感謝する。本研究はJSPS科研費24700100の助成を受けたものである。

参考文献

- 1) IDPF, "EPUB 3—International Digital Publishing Forum", <http://idpf.org/epub/30>
- 2) 高木利弘, "電子書籍ビジネス調査報告書", 三橋昭和, インプレスR&D, 東京(2009).
- 3) 株式会社コミックアニメーション, "コミックアニメーション「電子出版時代の「マンガ」の新しい「カタチ」を模索!", <http://www.comicanimation.jp/index.html>, 2013年2月26日参照.
- 4) P. Viola, M. J. Jones: "Robust Real-Time Face Detection", International Journal of Computer Vision, Vol. 57, No. 2, pp. 137-154 (2004).
- 5) D. G. Lowe: "Object Recognition from Local Scale-Invariant Features", Proc. of International Conference on Computer Vision, pp. 1150-1157 (1999).
- 6) N. Dalal, B. Triggs: "Histograms of Oriented Gradients for Human Detection", Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 886-893 (2005).
- 7) C. Cortes, V. N. Vapnik: "Support-Vector Networks", Machine Learning, Vol. 20, pp. 273-297 (1995).
- 8) K. Hoashi, C. Ono, D. Ishii, and H. Watanabe: "Automatic Preview Generation of Comic Episodes for Digitized Comic Search", Proc. of the 19th ACM International Conference on Multimedia, pp. 1489-1492 (2011).
- 9) 野中俊一郎, 沢野哲也, 羽田典久: "コミックスキャン画像からの自動コマ検出を可能とする画像処理技術「GT-Scan」の開発", 富士フィルム研究報告, 第57号, pp. 46-49 (2012).
- 10) 石井大祐, 渡辺 裕: "マンガからの自動キャラクター位置検

出に関する検討", 情報処理学会AVM研究会研究報告, Vol. 2012-AVM76, No. 2, pp. 1-5 (2012).

- 11) D. Ishii, T. Yamazaki, H. Watanabe: "Multi Size Eye Detection on Digitized Comic Image", Proc. of IIEEJ 3rd Image Electronics and Visual Computing Workshop, 1P-4 (2012).
- 12) 新井俊宏, 松井勇佑, 相澤清晴: "漫画画像からの顔検出", 2012電子情報通信学会総合大会, d-12-67, pp. 161 (2012).
- 13) K. Takayama, H. Johan, T. Nishita: "Face detection and face recognition of cartoon characters using feature extraction", IIEEJ 3rd Image Electronics and Visual Computing Workshop, 1C-5 (2012).
- 14) C. C. Chang, C. J. Lin: "LIBSVM: A Library for Support Vector Machines", ACM Trans. on Intelligent Systems and Technology, Vol. 2, No. 3, pp. 27:1-27:27 (2011).
- 15) opencv.org, "OpenCV", <http://opencv.org>.
- 16) 木野 陽: ベリーベリークリームショコラふたつのベリー, (2010).

(2013年2月8日受付)

(2013年6月27日再受付)



石井 大祐 (正会員)

2006年東京理科大学理学部第1部応用物理学卒業, 2008年早稲田大学大学院国際情報通信研究科修士課程修了。現在同研究科助手。修士。画像処理および画像解析に関する研究に従事。画像電子学会, 電子情報通信学会, 情報処理学会, IEEE各会員。



渡辺 裕 (正会員)

1980年北大・工・電子卒。1985年同大学院博士課程修了。同年NTT入社。NTT研究所において画像の高効率符号化及びMPEG標準化に従事。2000年早稲田大学大学院国際情報通信研究科教授。マルチメディアの符号化及び配信に関する研究に興味を持つ。工博。IEEE, 電子情報通信学会, 映像情報メディア学会, 情報処理学会, 画像電子学会各会員。